Classifying Smart Personal Assistants: An Empirical Cluster Analysis

Robin Knote
University of Kassel
robin.knote@uni-kassel.de

Andreas Janson
University of Kassel
janson@uni-kassel.de

Matthias Söllner
University of St.Gallen and
University of Kassel
matthias.soellner@unisg.ch

Jan Marco Leimeister
University of St.Gallen and
University of Kassel
janmarco.leimeister@unisg.ch

Abstract

The digital age has yielded systems that increasingly reduce the complexity of our everyday lives. As such, smart personal assistants such as Amazon's Alexa or Apple's Siri combine the comfort of intuitive natural language interaction with the utility of personalized and situation-dependent information and service provision. However, research on SPAs is becoming increasingly complex and opaque. To reduce complexity, this paper introduces a classification system for SPAs. Based on a systematic literature review, a cluster analysis reveals five SPA archetypes: Adaptive Voice (Vision) Assistants, Chatbot Assistants, Embodied Virtual Assistants, Passive Pervasive Assistants, and Natural Conversation Assistants.

1. Introduction

In recent years, technical progress has brought us systems that increasingly reduce the complexity of our everyday lives. Thereby, smart personal assistants (SPAs), defined as systems that use *"input such as the user's voice […] and contextual information to provide assistance by answering questions in natural language, making recommendations and performing actions"* [4, p. 223], have just conquered a broad consumer market. Recent forecasts predict the worldwide user count for SPAs such as Amazon Alexa, Apple's Siri or Microsoft Cortana to increase from 390 million in 2015 to 1.8 billion in 2021, which results in 2.3 billion USD average sales growth per year [33]. These systems' success story is mainly because digital assistants combine the comfort of intuitive natural language interaction with the utility of personalized and situation-dependent information and service provision. In practice, SPAs unfold their potential in various forms and contexts [8], such as on smartphones [38], in smart home environments [11], in cars [5], in service encounters [43], or as support for elderly or impaired people [11].

However, prominent examples such as those mentioned above represent SPAs that are explicitly developed for a broad consumer market. They thus are only the tip of the iceberg. Since the idea of information systems (IS) that pervasively assist humans in conducting certain tasks is by far not new, numerous efforts were made in IS, computer science and human-computer-interaction research to develop SPAs as previously defined. Simultaneously, research and practice has often neglected to 'stand on the shoulders of giants' by building up on each other's work. This has led to a partly overlapping diversity of concepts and terms for the developed artifact. For example, while many scholars entitle their SPA as a conversational agent, others would differ between mainly text-based and voice-based systems. Still others would label the text-based SPA as chatbot and the voice-based SPA as smart speaker. This example shows, that the range of possible terms for different types of SPAs differ heavily due to lacking conceptual clarity. The interchangeable use of terms has also been observed by other scholars [e.g., 8].

We, however, argue that conceptual clarity is highly important, not only for a correct categorization of SPAs to a higher-order group. It is also important for finding similarities and differences between systems, identifying design principles, recurring requirements and design practices (i.e., patterns) and, finally, reline future research and practice with a reliable structure to allocate SPA-related work. Therefore, this paper offers a classification approach for SPAs. Based on an exhaustive literature review, we derived design characteristics of 115 SPAs that were developed within a research project or for commercial

URI: https://hdl.handle.net/10125/59642
ISBN: 978-0-9981331-2-6

purposes. We further performed a k-means cluster analysis to yield groups of SPAs which, according to the design characteristics, have a high internal homogeneity (i.e., most similar items are within one cluster) and a high external heterogeneity (i.e., each cluster is highly distinctive to other clusters). An analysis of the clusters, their similarities and differences, resulted in archetypes of SPAs, which are defined by the most expressive design characteristics of each cluster. We thus aim to contribute to research by providing a classification for SPAs that aid future SPA research to yield more specific and meaningful contributions. We further contribute to practice by showing design differences between the various SPA types which may influence development decisions.

The remainder of the paper is structured as follows. Section 2 provides background information about SPAs to establish a shared understanding. We describe our methodology in section 3. In section 4 we present the results of our literature review and cluster analysis. Those are briefly discussed in section 5. The paper concludes with a short outlook.

2. Background

Although SPAs have just recently gained success on the consumer market, personal assistance provided by information systems (IS) is not a novel research topic at all. In the past, research in the field of artificial intelligence (AI) and focused on expert systems in relatively limited domains [18]. However, the advent of technical evolutions, such as cloud-service infrastructure, natural language processing, semantic reasoning, voice recognition and voice synthesis paved the way for modern SPAs such as Apple's Siri, Microsoft's Cortana, Samsung's Bixby, Amazon's Alexa, Google's Google Assistant and also chatbots in the service encounter. These smart service systems interact with the user via natural language and offer many opportunities of service and information provision to reduce effort and complexity of users' everyday tasks [8].

However, a general definition for SPAs (or respective synonyms) up until now is missing. A broad definition approach has already been conducted by Baber [4, p. 223] who considered an SPA to be *"an application that uses input such as the user's voice… and contextual information to provide assistance by answering questions in natural language, making recommendations and performing actions"*. More technical definitions stem from the field of computer science (CS) and draw on the term agent to describe SPAs. For example, Fuckner et al. [12, p. 89] describe an SPA as a "*specialized intelligent artificial agent that helps users to do their activities*" as an "*intermediary between humans and other agents in a multiagent environment.*" The term 'agent' aims to point out that the SPA as an autonomous entity is capable of perceiving and taking actions within its environment to achieve a certain goal [27], namely to assist the user conducting a specific task. Further, the SPA as an agent (e.g., Alexa) is able to interact with other agents, such as technical agents (e.g., a smart fridge) and human agents (users). The multi-agent concept also encompasses a layer view. Therein, an SPA consists of different layers, each conducting a specific sub-task (e.g., interface agent, interaction agent, transaction agent). For example, the user interacts with the interface agent which delegates more specific tasks to other types of agents [12]. In this context, the SPA serves as single, ubiquitous and easy-to-access entry point to a smart service infrastructure

The main purpose of SPAs is to enhance the user's perception, cognition and/or action abilities [16]. From a sociotechnical perspective and compared to other classes of information systems, the novelty of SPAs lies in two major aspects: the way how users *interact* with the device as well as the assistant's knowledgeability and human-like behavior, often summarized as artificial *intelligence* [21, 27]. Maedche et al. [21] suggest a classification of user assistance systems based on two dimensions, which we will use later on in this paper for cluster analysis: (1) the degree of intelligence of the system and (2) the degree of interaction implemented by the system. Advanced user assistance systems combine both intelligence and interaction to anticipate future situations and proactively adapt their assistance. From a service science perspective, SPAs can be considered agents in a broader smart service system [23].

3. Method

As a foundation for our cluster analysis, we conducted a systematic literature review[1] [39, 40] to identify SPAs developed for research and for commercial purposes. In detail, we first performed an open database search among *AISeL, IEEE Xplore, ACM DL, EBSCO Business Source Premier, ScienceDirect, ProQuest* and *Google Scholar* using the keywords *"smart assistant" OR "conversational agent" OR "virtual assistant" OR "assistance system" OR "personal assistant").* The search phase was adapted to fit databases' syntactic requirements and the

[1] A concept matrix that provides a detailed list of the reviewed articles and relations to our results is available at: http://downloads.wi-kassel.de/rkn/HICSS19/Knote_et_al_Appendix.pdf

search was limited to title, abstract, keywords and a publication period from 2000 to date. The initial open database search revealed 2802 hits. In order to reduce the results to manageable amount, we first screened and later thoroughly examined the literature regarding fit to the purpose of our study. Therefore, papers should either focus on conceptualizing or developing an SPA in parts or as a whole. In the 185 remaining papers, 83 SPAs could be identified that were developed as part of a research endeavor. We further reviewed the product websites of SPAs developed for commercial purposes (e.g., Amazon's various Echo devices) and included them to our data set. Altogether, we reviewed 115 SPAs to inductively derive design characteristics of *interaction* and *intelligence* for this class of systems. All reviewed systems were further assigned to these design characteristics by three independent researchers according to the design characteristics' definitions. This assertion procedure results in binary vectors for each SPA so that '1' indicates that the SPA obtains this design characteristic and '0' that it does not.

After all SPAs were assigned design characteristics, we performed a cluster analysis using k-means clustering in RStudio. The goal of a cluster analysis is to form groups of objects so that similar objects are in the same group and objects in different groups are as dissimilar as possible [17]. K-means, as one of the most prominent and efficient clustering algorithms, builds a previously defined number of k clusters from a set of similar objects. It therefore iteratively goes through several rounds of optimization until each object is closer to the centroid of the own group than that of any other group [19]. Since defining the number of clusters is a challenging task [1], cluster analysis is usually a two-step approach. First, we identified the optimal amount of clusters applying gap statistics, which can be used for any clustering algorithm to compare the change in within-cluster dispersion with that expected under an appropriate reference null distribution, i.e. a distribution with no obvious clustering [35]. We computed gap statistics for k-means clustering of our data set (i.e., a data frame of 115 binary vectors) using the *NbClust* und *factoextra* libraries. Gap statistic indicates that five clusters are the optimal amount for clustering our data set via k-means. The scree plot for the gap statistics is shown in figure 1. Hence, the clustering algorithm computes with a cluster count of k = 5. Since k-means clustering starts with k randomly selected centroids, we first set a seed for R's random number generator via set.seed(123). This is especially important to yield reproducible results for scholarly purposes. We then performed the actual k-means clustering with k = 5, Euclidean distance measure, which is suitable for binary vectors, and 25 different random starting assignments from which R selected the best result corresponding to the one with the lowest within-cluster variance.

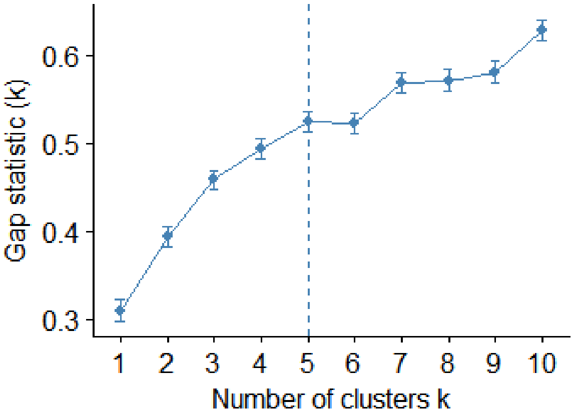

Figure 1. Scree plot for the optimal number of clusters according to gap statistics

Afterwards, we manually assessed the clusters for their meaning according to the cluster center values for each design characteristic. Hence, the five clusters represent archetypes of SPAs which can be defined by their predominant design characteristics.

4. Results

In the following, we present the results of our study, namely typical SPA design characteristics from the literature and archetypes (i.e., clusters) of SPAs.

4.1 SPA Design Characteristics

Based on intelligence and interaction as salient SPA design factors, we inductively narrowed these high-level constructs down to concrete design attributes. Thereby, we payed special attention to formulate design characteristics so that they are mutually and collectively exclusive and that each design characteristic is obtained by at least one SPA. We found 31 design characteristics (*italic)* and grouped them into 10 dimensions (**bold**). Characteristics that specify the degree of interaction are:

Communication mode: the primary way(s) a user communicates with an SPA and vice-versa. Communication is either based on user-entered and/or SPA generated *text* [28], user's and/or synthesized *voice* [41], *vision* sensors, cameras and generated animations [16], a combination of *voice and vision* (including text) [15], or *observational* sensing and/or unconscious acting (i.e., assistance is not inevitably augmentable for the user; 7).

Direction of explicit interaction: comprises *user-to-system* interaction [6], *system-to-user* interaction [29] and *bidirectional* interaction [37]. User-to-system interaction means that the user provides input which is intentionally and consciously directed towards the SPA. The system's response may be unconscious for the user. System-to-user interaction means that an SPA addresses the conscious mind to create a change in the environment that the user cannot avoid consciously perceiving [16]. In this case the user does not put an explicit request upfront but rather receives the result of the SPA's ability to passively observe and make sense of context information. Bidirectional interaction means that it delivers services in communicational exchange.

Query input: the way in which users formulate requests towards the SPA. Requests can either be predefined *formal prompts* that users must know to trigger a desired action [37], *natural language* requests [31] or accumulations of *sensor data* which, from a user perspective, is often collected unconsciously [9].

Response output: the way in which an SPA formulates responses to user requests. An SPA provides *visual* output if it responds via text, images, videos, an avatar or by any combination of the aforementioned [25]. *Voice* output refers to responses via synthesized speech as it is common for most commercial SPAs currently available [30]. SPAs that combine visual and verbal responses, such as smart speakers with an integrated screen, are classified as *voice and vision* [18].

Action: An SPA's capabilities to execute services based on query input. One can broadly distinguish between the general ability to, for example, play music, set alarms or control smart household objects as part of a larger smart service system *(service execution)* [15] and 'simple' functionality such as question answering and information retrieval *(no service execution)* [31].

Design characteristics to specify the degree of SPA intelligence are:

Assistance domain: determines both the functionalities and the knowledge models (i.e., semantic models like ontologies) that must be implemented to provide appropriate assistance for a given context. An SPA may either provide *general assistance* like retrieving information, searching on the web or playing music [28], or *specific assistance* for certain complex tasks [18, 31] or to a dedicated user group [16].

Accepted commands: Provide control over the SPA's behavior. The simplest form is *manual data entry* [7], followed by *simple commands* such as "send email to Jeff" [41] and *compound commands* such as "every day at 6am get the latest weather and send it via email to Jeff" [6]. However, some SPAs *do not* offer the user the ability to control system behavior [38].

Adaptivity: the system's ability to learn by interpreting (usually a rich amount of) data and adapt assistance services accordingly. Examples are the improvement of speech recognition [3] or tailored interaction for different users over time [2]. An SPA is characterized to have either *static behavior*, if service provision is not reflected and revised against data [14], or *adaptive behavior* if assistance is a function of context or prior assistance [6].

Collective intelligence: the ability to learn, to understand, and to adapt to an environment by using the knowledge of the user crowd [20]. SPAs may leverage the potentials of collective intelligence to improve machine learning algorithms and, thus, increase service quality. For example, the analysis of many users' natural language utterances may lead to a steeper learning curve for speech recognition algorithms since adaptivity is based on a large and heterogenous data set. Hence, individual SPA users may benefit from *crowd engagement* [6]. However, some SPAs *do not* leverage the potentials of crowd engagement [30].

Embodiment: the aspiration to present the user a clearly identifiable counterpart who provides personal assistance. In SPAs, this is mostly accomplished through anthropomorphism, "a conscious mechanism wherein people infer that a non-human entity has human-like characteristics and warrants human-like treatment" [26, p. 2854]. Embodied or anthropomorphic design is usually applied to provide a shared common ground, represent an authentic entity, combine verbal and non-verbal communication and align minds by being interesting, creative and humorous [22]. In practice, embodiment is accomplished by *virtual characters*, i.e., avatars [10, 24], a (often human-like) computer *voice* [36] or a *combination* of both [44]. However, some SPAs *do not* use embodiment at all [38].

The second column of table 1 shows the distribution of the 115 SPAs over intelligence and interaction design characteristics.

4.2 SPA Archetypes

k-means clustering reveals five distinctive groups of objects. Columns 3 to 7 of table 1 show cluster means for each design characteristic. We further manually reviewed the clusters regarding predominant design characteristics (i.e. high cluster means) and representative objects to suggest five SPA archetypes. A list of SPAs and respective cluster assertions is provided in the appendix.

Table 1. SPA distribution over design characteristics and cluster means

Characteristics	*SPAs*	*C1 (26)*	*C2 (19)*	*C3 (38)*	*C4 (15)*	*C5 (17)*
communication mode						
text	18	3,8%	68,4%	5,3%	6,7%	5,9%
voice	23	26,9%	0,0%	5,3%	6,7%	76,5%
vision	3	0,0%	5,3%	0,0%	13,3%	0,0%
text and vision	6	0,0%	10,5%	2,6%	13,3%	5,9%
voice and vision	57	69,2%	15,8%	86,8%	6,7%	11,8%
passive / observational	8	0,0%	0,0%	0,0%	53,3%	0,0%
direction of explicit interaction						
user-to-system	4	3,8%	5,3%	2,6%	0,0%	5,9%
system-to-user	18	0,0%	5,3%	10,5%	86,7%	0,0%
bidirectional	93	96,2%	89,5%	86,8%	13,3%	94,1%
query input						
formal prompts	12	3,8%	26,3%	10,5%	0,0%	11,8%
natural language	83	96,2%	68,4%	78,9%	0,0%	88,2%
sensor data	20	0,0%	5,3%	10,5%	100,0%	0,0%
response output						
vision	35	19,2%	89,5%	0,0%	80,0%	5,9%
voice	20	23,1%	5,3%	5,3%	13,3%	52,9%
voice and vision	60	57,7%	5,3%	94,7%	6,7%	41,2%
action						
no service execution	65	0,0%	94,7%	94,7%	53,3%	17,6%
service execution	50	100,0%	5,3%	5,3%	46,7%	82,4%
assistance domain						
general	45	96,2%	26,3%	10,5%	13,3%	52,9%
specific	70	3,8%	73,7%	89,5%	86,7%	47,1%
accepted commands						
none	50	0,0%	47,4%	68,4%	86,7%	11,8%
manual data entry	17	0,0%	47,4%	13,2%	13,3%	5,9%
primitive commands	36	96,2%	5,3%	10,5%	0,0%	35,3%
compound commands	12	3,8%	0,0%	7,9%	0,0%	47,1%
adaptivity						
static behavior	64	0,0%	68,4%	68,4%	86,7%	70,6%
adaptive behavior	51	100,0%	31,6%	31,6%	13,3%	29,4%
collective intelligence						
no crowd engagement	93	19,2%	100,0%	97,4%	100,0%	100,0%
crowd engagement	22	80,8%	0,0%	2,6%	0,0%	0,0%
embodiment						
none	30	23,1%	47,4%	0,0%	80,0%	17,6%
virtual character	14	3,8%	52,6%	0,0%	13,3%	5,9%
artificial voice	28	65,4%	0,0%	2,6%	6,7%	52,9%
virtual character with voice	43	7,7%	0,0%	97,4%	0,0%	23,5%

Cluster 1 – Adaptive Voice (Vision) Assistants: The first group contains SPAs that assist users mainly via speech and, optionally, also via optical sensors and visual output on a screen. Although most objects in this group combine speech control with visual interaction, such as gesture control over integrated cameras or supplemental on-screen information, speech currently remains the predominant interaction mode. Therefore, these systems are capable of both understanding and responding in natural language and execute (also third-party) services upon user requests. The vast majority of type 1 SPAs obtains knowledge models for general purposes, such as controlling smart household gadgets, retrieving mails or adding calendar entries. These knowledge models, however, are adaptive as they evolve over longer usage periods and, thus, provide higher service quality when used regularly. This mostly concerns the natural language processing behavior, which means that human utterances are understood and interpreted more correctly the more often the SPA is used. Thereby, adaptivity usually leverages collective intelligence. Speech and usage data of a broad range of users is recorded and stored in large data centers (or dedicated cloud environments) and processed to improve service quality of the SPA. Hence, individual users profit from experiences and interactions of other members within the user crowd. Further, since speech is the predominant interaction mode, most type 1 SPAs are embodied via a (usually human-like) computer-generated voice. Due to their advanced adaptivity and predominant interaction modes, we entitle this group of SPAs Adaptive Voice (Vision) Assistants. Prominent examples of this SPA class are the Amazon devices running Alexa, Apple’s Siri, Google’s Assistant, Microsoft’s Cortana and Samsung’s S Voice and Bixby. It should be noticed that nearly all consumer-oriented SPAs in this investigation belong to this class. This is because the high user count of commercial systems makes it easier to leverage collective intelligence potentials for system adaptivity and service quality optimization.

Cluster 2 – Chatbot Assistants: The second cluster contains SPAs which mainly rely on text chat interaction to provide assistance services. This especially comprises chatbots, text-based conversational agents that are able to react to user input based on semantic text analysis. Such systems are increasingly employed in first-level support service encounters as they are able to answer frequently asked questions and guide users through support processes. To simplify their handling, most chatbot assistants are capable of interpreting natural language and respond accordingly. Since the exchange is text-based, interaction is conducted over screens which, however, may also show supplemental information, such as images or videos according to the user’s request. Chatbot assistants usually encompass rather specific (domain) knowledge and are used to present information rather than to execute (third-party) services. All type 2 SPAs in our study are implemented as rather closed systems that do not leverage the wisdom of the crowd. Knowledge models, however, may adapt to individual user’s usage patterns. While a great number of chatbot SPAs provide fields for text in-and output only, more than half are designed to enhance the user experience via virtual characters, such as avatars. A representative example for this class of systems is MentorChat, a configurable text-based agent for collaborative learning [32].

Cluster 3 – Embodied Virtual Assistants: The largest class comprises SPAs which are embodied by, often human-like, virtual assistants. This is accomplished by both speech and visual output. Systems are mainly screen-based to present a virtual character (or avatar) with natural language speech, mimics and gestures to provide familiar interaction. Often, these assistants are designed for a special purpose such as e-learning, which is the biggest domain for type 3 SPAs. In the comparably rare cases that users have any control over the system’s behavior, type 3 SPAs mainly accept manual data entry of values or simple commands (e.g., for adjusting severity levels in e-learning). However, about one third of these systems is able to adjust to user’s preferences or behavior autonomously. A much smaller amount therefore leverages collective intelligence since most adaptive systems focus on the individual user and do not infer actions based on similar behavioral patterns of crowd members. The aim of type 3 SPAs is to enhance user interaction by seamlessly transferring prior human-to-human activities, such as tutoring, to the virtual world while remaining benefits of human interaction, such as empathy, humor and learner context [34]. As mentioned earlier, anthropomorphism is suggested to be efficient for increasing SPA acceptance and, thus, positively influence outcomes of system use (e.g., improved learning curve). AutoTutor [13] and Victor, the virtual tutor [14] are both representative examples.

Cluster 4 – Passive Pervasive Assistants: While all prior SPA types focus on bidirectional and explicit exchange, type 4 SPAs are designed to be as unobtrusive as possible. This means that users have few interactions with the system itself while it collects data from (usually multiple different) sensors and infers and recommends suitable action. In other words, manual user input is not required for the SPA to provide relevant information and advice. Hence, explicit interaction is usually initiated by the system which passively observes the user's tasks and context. Assistance is thereby mainly provided via screen output. In addition, almost half of all SPAs under investigation autonomously perform actions as a reaction on sensed trigger events (e.g., changing the color of the lights according to the user's mood). Most type 4 SPAs offer assistance for specialized purposes, such as cooking or sightseeing. Underlying knowledge models are seldom adaptive to observed context or task patterns and none of the systems under investigation uses crowd-generated data for service quality improvements. Since passive pervasive assistants are designed to not actively disturb the user's conscious mind, they usually are not embodied at all. Rather, the physical environment and the SPA with all its sensors and actuators should seamlessly conflate into a digitally enhanced experience. One representative example for such an enhancement is MimiCook, a ubiquitous cooking assistant which is integrated into the physical kitchen environment [29].

Cluster 5 – Natural Conversation Assistants: This class of SPAs can also be considered assistant for the 'next generation of service encounters'. Focusing on speech interaction, type 5 SPAs aim to increase the similarity to human-to-human natural language interaction. They thus encompass more sophisticated speech recognition and spoken language understanding capabilities than any other class. Hence, they are more likely to understand and being controlled by complex compound than type 1 SPAs. The primary design goal is to imitate human natural language interaction to provide a most natural and familiar interaction experience. This requires the underlying linguistic model to not only respond to human utterances correctly but also to work with fillers such as "ah", "um" or pauses. Like type 2 SPAs, natural conversation assistants may be used as agents for first level support, e.g., as single point of contact in a call center. While this would require a high level of adaptivity and, eventually, collective intelligence, most current type 5 SPAs show rather static behavior. One prominent example for this class, is Google Duplex, a conversational agent which may behave confusingly similar to a human agent.

5. Discussion

In the previous section, we elaborated on the results of our systematic literature review and our k-means cluster analysis. Based on advanced intelligence and interaction as salient factors for SPAs, our review of 115 SPAs inductively revealed 31 design characteristics, which we grouped into 10 dimensions. We further conducted a cluster analysis and provided a classification approach for SPAs. Classifications are fundamental to provide a structure for further research and development activities as it helps understanding the science behind design principles of observed artifacts [42]. Especially in the domain of SPA research, we observe that research streams become increasingly opaque and diverse. This is mainly because of the fragmented use of heterogenous terms and their interpretations which impedes the search for unified definitions. With our literature review and clustering approach, we made a step towards solving this issue by fostering conceptual clarity and providing a framework of SPA archetypes. Future conceptual, empirical or design-oriented research may now contribute to certain types of SPAs more clearly. Furthermore, we contribute to practice by providing baselines for SPA development. However, our research does not come without limitations. First, all results depend on our understanding and interpretation of the literature base. Although we have profound knowledge in the field of SPAs, there cannot be a guarantee for 'objective validity' of the clusters. We therefore highly encourage future research to challenge and enhance our classification system with different design characteristics. Second, k-means clustering has some weaknesses. For example, it assumes the researcher to define the optimal number of clusters in advance and is sensitive to outliers. While we could manage finding an appropriate number of clusters with gap statistics, future research should investigate the role and nature of objects that do not entirely fit in one of the clusters. Despite all limitations, we think that our results reveal a useful and valuable classification scheme, which, to the best of our knowledge, is the first of its kind.

6. Conclusion

This paper provides a classification framework for SPAs based on a systematic literature review and cluster analysis. As research on highly intelligent and interactive SPAs is still in its infancy, we hope to set a solid foundation for future conceptual, empirical or design-oriented endeavors.

7. Acknowledgements

This paper presents research that was conducted in context of the projects "AnEkA" (project number: 348084924), funded by the German Research Foundation (DFG) and "KoLeArn" (funding number: 01BE17008A, managed by the DLR Project Management Agency), funded by the German Federal Ministry of Education and Research (BMBF). The responsibility for the content of this publication remains with the authors.

8. References

[1] Anderberg, M.R., Cluster analysis for applications, Office of the Assistant for Study Support Kirtland AFB N MEX, 1973.

[2] Armentano, M., D. Godoy, and A. Amandi, "Personal assistants: Direct manipulation vs. mixed initiative interfaces", International Journal of Human-Computer Studies, 64(1), 2006, pp. 27–35.

[3] Arsikere, H. and S. Garimella, "Robust Online i-Vectors for Unsupervised Adaptation of DNN Acoustic Models: A Study in the Context of Digital Voice Assistants", in Interspeech 2017. 2017. ISCA.

[4] Baber, C., "Developing interactive speech technology", Taylor & Francis, Inc, 1993.

[5] Bengler, K., K. Dietmayer, B. Farber, M. Maurer, C. Stiller, and H. Winner, "Three Decades of Driver Assistance Systems: Review and Future Perspectives", IEEE Intelligent Transportation Systems Magazine, 6(4), 2014, pp. 6–22.

[6] Campagna, G., R. Ramesh, S. Xu, M. Fischer, and M.S. Lam, "Almond: The Architecture of an Open, Crowdsourced, Privacy-Preserving, Programmable Virtual Assistant", International World Wide Web Conferences Steering Committee, 2017.

[7] Chen, C.-C., T.-C. Huang, J.J. Park, H.-H. Tseng, and N.Y. Yen, "A smart assistant toward product-awareness shopping", Personal and Ubiquitous Computing, 18(2), 2014, pp. 339–349.

[8] Cowan, B.R., N. Pantidi, D. Coyle, K. Morrissey, P. Clarke, S. Al-Shehri, D. Earley, and N. Bandeira, ""What can i help you with?": Infrequent Users’ Experiences of Intelligent Personal Assistants", in Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services - MobileHCI '17, Vienna, Austria. 2017.

[9] Czibula, G., A.-M. Guran, I.G. Czibula, and G.S. Cojocar, "IPA - An Intelligent Personal Assistant Agent for Task Performance Support", in Proceedings of the 5th IEEE International Conference on Intelligent Computer Communication and Processing. 2009.

[10] Ernst, S.-J., A. Janson, M. Söllner, and J.M. Leimeister, "It's about Understanding Each Other's Culture - Improving the Outcomes of Mobile Learning by Avoiding Culture Conflicts", in ICIS 2016 Proceedings. 2016.

[11] Fernando, N., F.T.C. Tan, R. Vasa, K. Mouzaki, and I. Aitken, "Examining Digital Assisted Living: Towards a Case Study of Smart Homes for the Elderly", in Proceedings of the 24th European Conference on Information Systems (ECIS). 2016: Istanbul, Turkey.

[12] Fuckner, M., J.-P. Barthes, and E.E. Scalabrin, "Using a personal assistant for exploiting service interfaces", in Proceedings of the 2014 IEEE 18th International Conference on Computer Supported Cooperative Work in Design (CSCWD), Hsinchu, Taiwan. 2014.

[13] Graesser, A.C., P. Chipman, B.C. Haynes, and A. Olney, "AutoTutor: An intelligent tutoring system with mixed-initiative dialogue", IEEE Transactions on Education, 48(4), 2005, pp. 612–618.

[14] Grujic, Z., B. Kovaeic, and I.S. Pandzic, "Building Victor-A virtual affective tutor", in 10th International Conference on Telecommunications (ConTEL) 2009. 2009.

[15] Hauswald, J., T. Mudge, V. Petrucci, L. Tang, J. Mars, M.A. Laurenzano, Y. Zhang, H. Yang, Y. Kang, C. Li, A. Rovinski, A. Khurana, and R.G. Dreslinski, "Designing Future Warehouse-Scale Computers for Sirius, an End-to-End Voice and Vision Personal Assistant", ACM Transactions on Computer Systems, 34(1), 2016, pp. 1–32.

[16] Jalaliniya, S. and T. Pederson, "Designing Wearable Personal Assistants for Surgeons: An Egocentric Approach", IEEE Pervasive Computing, 14(3), 2015, pp. 22–31.

[17] Kaufman, L. and P.J. Rousseeuw, Finding groups in data: An introduction to cluster analysis, John Wiley & Sons, 2009.

[18] Kincaid, R. and G. Pollock, "Nicky: Toward a Virtual Assistant for Test and Measurement Instrument Recommendations", in 2017 IEEE 11th International Conference on Semantic Computing (ICSC). 2017.

[19] Landau, S., A handbook of statistical analyses using SPSS, CRC, 2004.

[20] Leimeister, J.M., "Collective Intelligence", Business & Information Systems Engineering, 2(4), 2010, pp. 245–248.

[21] Maedche, A., S. Morana, S. Schacht, D. Werth, and J. Krumeich, "Advanced User Assistance Systems", Business & Information Systems Engineering, 58(5), 2016, pp. 367–370.

[22] McKeown, G., "Turing's menagerie: Talking lions, virtual bats, electric sheep and analogical peacocks: Common ground and common interest are necessary components of engagement", in International Conference on Affective Computing and Intelligent Interaction (ACII) 2015, Xi'an, China. 2015.

[23] Medina-Borja, A., "Editorial Column—Smart Things as Service Providers: A Call for Convergence of Disciplines to Build a Research Agenda for the Service Systems of the Future", Service Science, 7(1), 2015, pp. ii–v.

[24] Ochs, M., C. Pelachaud, and G. Mckeown, "A User Perception--Based Approach to Create Smiling Embodied Conversational Agents", ACM Transactions on Interactive Intelligent Systems, 7(1), 2017, pp. 1–33.

[25] Onorati, T., A. Malizia, K.A. Olsen, P. Diaz, and I. Aedo, "I feel lucky: An automated personal assistant for smartphones", AVI '12 Proceedings of the International Working Conference on Advanced Visual Interfaces, Capri Island, Italy, 2012, pp. 328–331.

[26] Purington, A., J.G. Taft, S. Sannon, N.N. Bazarova, and S.H. Taylor, ""Alexa is my new BFF"", in Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems, Denver, Colorado, USA. 2017. ACM Press: New York, New York, USA.

[27] Russell, S.J. and P. Norvig, Künstliche Intelligenz: Ein moderner Ansatz, 3rd edn., Pearson, München, 2012.

[28] Sansonnet, J.P., D.W. Correa, P. Jaques, A. Braffort, and C. Verrecchia, "Developing web fully-integrated conversational assistant agents", Proceedings of the 2012 ACM Research in Applied Computation Symposium, 2012, pp. 14–19.

[29] Sato, A., K. Watanabe, and J. Rekimoto, "MimiCook: A cooking assistant system with situated guidance", Proceedings of the 8th International Conference on Tangible, Embedded and Embodied Interaction, 2014, pp. 121–124.

[30] Schmeil, A. and W. Broll, "MARA - A Mobile Augmented Reality-Based Virtual Assistant", in IEEE Virtual Reality Conference, Charlotte, NC, USA. 2007.

[31] Sugawara, K., Y. Manabe, N. Shiratori, S.B. Yaala, C. Moulin, and J.-P.A. Barthes, "Conversation-based support for requirement definition by a Personal Design Assistant", in Proceedings of the 10th IEEE International Conference on Cognitive Informatics & Cognitive Computing, Banff, AB, Canada. 2011.

[32] Tegos, S., S. Demetriadis, and A. Karakostas, "MentorChat: Introducing a Configurable Conversational Agent as a Tool for Adaptive Online Collaboration Support", in Proceedings of the 2011 Panhellenic Conference on Informatics (PCI), Kastoria, Greece. 2011.

[33] https://www.tractica.com/newsroom/press-releases/the-virtual-digital-assistant-market-will-reach-15-8-billion-worldwide-by-2021/, accessed 8-20-2017.

[34] Thiel de Gafenco, M., A. Janson, and T. Schneider, "KoLeArn – Smarte und kontextsensitive Aus- und Weiterbildung für die chinesische Industrie", in DeLFI 2018 - Die 16. E-Learning Fachtagung Informatik, D. Krömker and U. Schroeder, Editors. 2018. Gesellschaft für Informatik e.V: Bonn.

[35] Tibshirani, R., G. Walther, and T. Hastie, "Estimating the number of clusters in a data set via the gap statistic", Journal of the Royal Statistical Society: Series B, 63(Part 2), 2001, pp. 411–423.

[36] Trovato, G., J.G. Ramos, H. Azevedo, A. Moroni, S. Magossi, H. Ishii, R. Simmons, and A. Takanishi, "Designing a receptionist robot: Effect of voice and appearance on anthropomorphism", in 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN). 2015.

[37] Tsujino, K., S. Iizuka, Y. Nakashima, and Y. Isoda, "Speech Recognition and Spoken Language Understanding for Mobile Personal Assistants: A Case Study of "Shabette Concier"", in IEEE 14th International Conference on Mobile Data Management (MDM), 2013, Milan, Italy. 2013.

[38] Venkatesh, V., J.A. Aloysius, H. Hoehle, and S. Burton, "Design and Evaluation of Auto-ID Enabled Shopping Assistance Artifacts in Customer's Mobile Phones: Two Retail Store Laboratory Experiments", MIS Quarterly, 41(1), 2017, pp. 83–113.

[39] Vom Brocke, J., A. Simons, K. Riemer, B. Niehaves, R. Plattfaut, and A. Cleven, "Standing on the Shoulders of Giants: Challenges and Recommendations of Literature Search in Information Systems Research", Communications of the Association for Information Systems, 37(1), 2015.

[40] Webster, J. and R.T. Watson, "Analyzing the Past to Prepare for the Future: Writing a Literature Review", MIS Quarterly, 26(2), 2002, pp. xiii–xxiii.

[41] Weeratunga, A.M., S.A.U. Jayawardana, P.M.A.K. Hasindu, W.P.M. Prashan, and S. Thelijjagoda, "Project Nethra - an intelligent assistant for the visually disabled to interact with internet services", in IEEE 10th International Conference on Industrial and Information Systems (ICIIS), Peradeniya, Sri Lanka. 2015.

[42] Williams, K., S. Chatterjee, and M. Rossi, "Design of emerging digital services: A taxonomy", European journal of information systems, 17(5), 2008, pp. 505–517.

[43] Xu, A., Z. Liu, Y. Guo, V. Sinha, and R. Akkiraju, "A New Chatbot for Customer Service on Social Media", in Proceedings of the Annual CHI Conference on Human Factors in Computing Systems, Denver, Colorado, USA. 2017. ACM: New York, NY.

[44] Zoric, G., K. Smid, and I.S. Pandzic, "Automated gesturing for virtual characters: Speech-driven and text-driven approaches", in Image and Signal Processing and Analysis, 2005. ISPA 2005. Proceedings of the 4th International Symposium on. 2005.

9. Appendix: SPA Classification

Reference (SPA Name)*	*Cluster*
Campagna 2017 ("Almond")	1
Mihale-Wilson et al. 2017	1
Wang 2016 ("Duer")	1
Apple 2011 ("Siri")	1
Microsoft 2014 ("Cortana")	1
Google 2016 ("Google Assistant")	1
Samsung 2012 ("S Voice")	1
Nuance 2012 ("Nina")	1
BlackBerry 2014 ("BlackBerry Assistant")	1
Cognitive Code 2008 ("SILVIA")	1
Viv Labs 2016 ("Viv")	1
Nuance ("Dragon Go!")	1
Aido 2018 ("Aido")	1
Samsung 2017 ("Bixby")	1
Brainasoft 2015 ("Braina Virtual Assistant")	1
Amazon 2017 ("Echo Plus")	1
Amazon 2015 ("Echo Dot")	1
Amazon 2017 ("Echo Look")	1
Amazon 2017 ("Echo Show")	1
Amazon 2018 ("Echo Spot")	1
Amazon 2015 ("Tap")	1
Sonos 2017 ("Sonos One")	1
Lenovo 2018 ("Lenovo Smart Assistant")	1
Amazon 2014 ("Fire TV")	1
Amazon 2012 ("Fire 7-Tablet")	1
Amazon 2017 ("Dash Wand")	1
Tegos et al. 2011-2015 ("MentorChat")	2
Abdelkefi/ Kallel 2016 ("MobiSpeech")	2
Armento et al. 2006	2
Derrick/Ligon 2014 ("Pat")	2
Dybala et al. 2010 ("MAS Punda")	2
Fudholi et al. 2009	2
Hacker et al. 2009 ("xGECA")	2
Kerly et al. 2008 ("CALMsystem")	2
Latham et al. 2010 ("Oscar")	2
Niewiadomskia/Pelachaudb 2010	2
Pérez et al. 2016 ("E-VOX")	2
Perez-Marin/ Pascual-Nieto 2013 ("Shamael")	2
Schouten et al. 2017	2
Song et al. 2017	2
Sugawara et al. 2011 ("PDA")	2
van der Zwaan/ Dignum 2013 ("Robin")	2
Yoshii/ Nakajima 2015 ("Fairy Agent")	2
Green Jr. et al 1961 ("BASEBALL")	2
Weizenbaum 1966 ("ELIZA")	2
Graesser et al. 2005 ("AutoTutor")	3
Santos-Perez et al. 2013	3
Augello et al. 2008 ("Humorist Bot")	3
Ayedoun et al. 2015	3
Bickmore et al. 2013	3
Boukricha/ Wachsmuth 2011 ("EMMA")	3
Cassell, 2000 ("Rea")	3
Cavazza et al. 2010 ("HWYD Companion")	3
Datta/ Vijay 2010 ("Neel")	3
den Os et al. 2005	3
Doumanis/ Smith 2014	3
Gris et al. 2016 ("Young Merlin")	3
Grujic et al. 2009 ("Victor")	3
Hasegawa et al. 2014	3
Hayashi 2013	3
Hoque et al. 2013 ("MACH")	3
Huang et al. 2011	3
Hubal et al. 2008	3
Ishii et al. 2013	3
Kanaoka/ Mutlu 2015 ("Nao")	3
Kincaid/Pollock 2017 ("Nicky")	3
Krämer et al. 2013 ("Max")	3
Lisetti et al. 2013 ("ODVIC")	3
López et al. 2008	3
Miyake/ Ito 2012	3
Moussa et al. 2010	3
Niculescu et al. 2014 ("SARA")	3
Nunamaker et al. 2011	3
Rudra et al. 2012 ("ESCAP")	3
Schmeil/Broll 2007 ("MARA")	3
Sing Goh et al. 2006 ("AINI")	3
Trinh et al. 2015 ("DynamicDuo")	3
Trovato et al. 2005; 2015 ("Ana" / "KOBIAN")	3
Wainer et al. 2014 ("KASPAR")	3
Wargnier et al. 2016 ("Louise")	3
Yang et al. 2017 ("Zara the Supergirl")	3
Zhang et al. 2017	3
Zia-ul-Haque et al. 2007	3
De Carolis et al. 2015 ("DIVA")	4
Sansonnet et al. 2012 ("DIVAlite")	4
Chen et al. 2014	4
Czibula et al. 2009 ("IPA Agent")	4
Imtiaz et al. 2014	4
Iwamura et al. 2014	4
Jalaliniya and Pederson 2015	4
Lakde/ Prasad 2015	4
Nam et al. 2016	4
Onorati et al. 2012 ("I feel Lucky")	4
Öyzurt et al. 2013 ("COGAS")	4
Santos et al. 2016	4
Sato et al. 2014 ("MimiCook")	4
Vales-Alonso et al. 2015 ("SAETA")	4
Xiahou/ Xing 2010 ("WTAS Framework")	4
Adam et al. 2010	5
Eismann et al. 2016	5
Garcia-Serrano et al. 2004 ("ADVICE Project")	5
Gnjatovi et al. 2012	5
Griol et al. 2003 ("DI@L-log")	5
Hauswald et al. 2016 ("Sirius")	5
Paraiso, Barthes 2006	5
Teixeira et al. 2014 ("PaeLife")	5
Tsujino et al. 2013 ("Shabette Concier")	5
Weeratunga et al. 2015 ("Nethra")	5
Woods/ Kaplan 1977 ("LUNAR")	5
Hey Athena 2016 ("Hey Athena")	5
SoundHound 2015 ("Hound")	5
Jibo 2017 ("Jibo")	5
Clarity Lab 2015 ("Lucida")	5
Mycroft AI 2018 ("Mycroft")	5
Google 2018 ("Google Duplex")	5

(*references omitted due to space limitations; for detailed references see concept matrix in the online appendix)